

Primene paralelnog računarstva

Primene paralelnog računarstva

- **Složena i obimna izračunavanja**
- **Obrada velike količine podataka** (oblasti nauke o podacima (engl. *data science*) i sistema za rad sa velikim količinama podataka (engl. *big data*))
 - Internet (veb pretraživači), genetika (DNK), medicinske i astronomske slike, senzori (engl. *Internet-of-Things*, IoT)
- **Simulacija**
- **Vizuelizacija**
- Svi prethodno opisani problemi zahtevaju znatne resurse za izračunavanja, vode nas do oblasti **računarstva visokih performansi** (engl. *High Performance Computing*)
 - Klasteri, data centri, super-računari
 - Programiranje super-računara: ssh protokol za komunikaciju, MPI + OpenMP/CUDA/OpenACC

Primene paralelnog računarstva

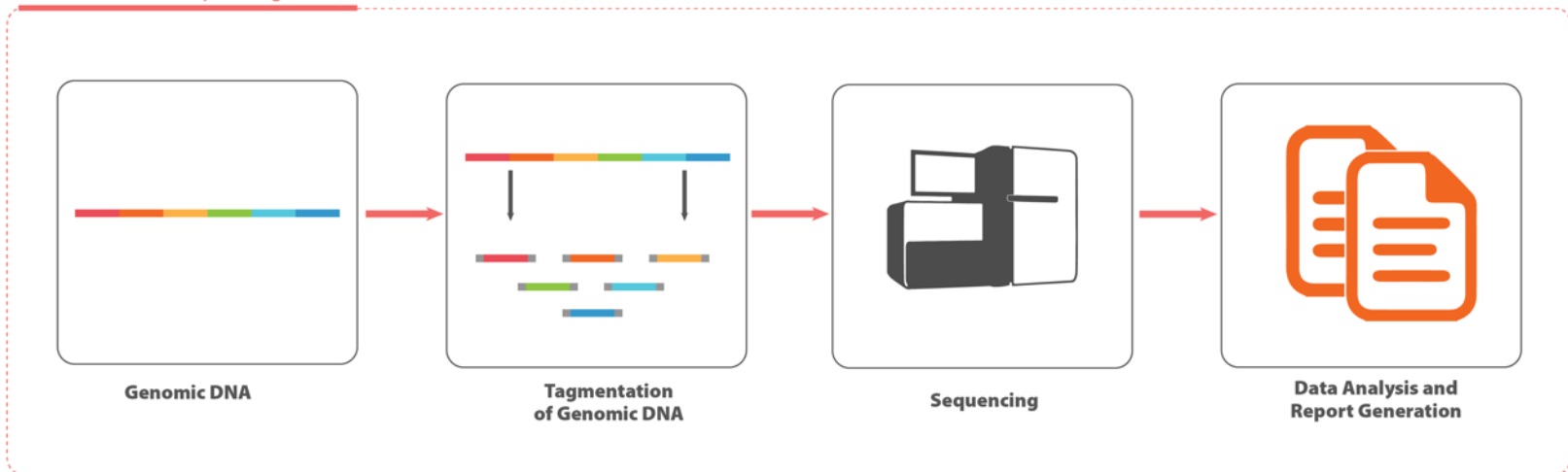
- **Naučna izračunavanja** (engl. *scientific computing*)
 - Klima i vremenska prognoza, genetika, astrofizika, okeanografija, seizmologija, farmacija, ekonomija i finansije, ...
- **Veštačka inteligencija** (engl. *artificial intelligence*)
 - Mašinsko učenje, duboke neuronske mreže, ...
- **Inženjerstvo**
 - Projektovanje vozila, letelica i mašina, razvoj novih materijala, ...
- **Energetika**
 - Obnovljivi izvori energije, istraživanje tla , ...
- **Medicina**
 - Obrada medicinskih slika, ...



Naučna izračunavanja – genetika

- **Sekvenciranje genoma** podrazumeva vrlo kompleksna izračunavanja (algoritmi za poklapanje, edit distance i rad sa stringovima), Human Genome Project doveo do sekvenciranja ljudskog genoma 2001.
- **Napredak paralelnog računarstva** omogućio snižavanje cene sekvenciranja
- Osnova za razvoj bioinformatike, novih lekova i **personalizovane medicine**

Whole Genome Sequencing (WGS)



Izvor: <https://www.abmgood.com/Whole-Genome-Sequencing-Service.html>

Naučna izračunavanja – ekonomija

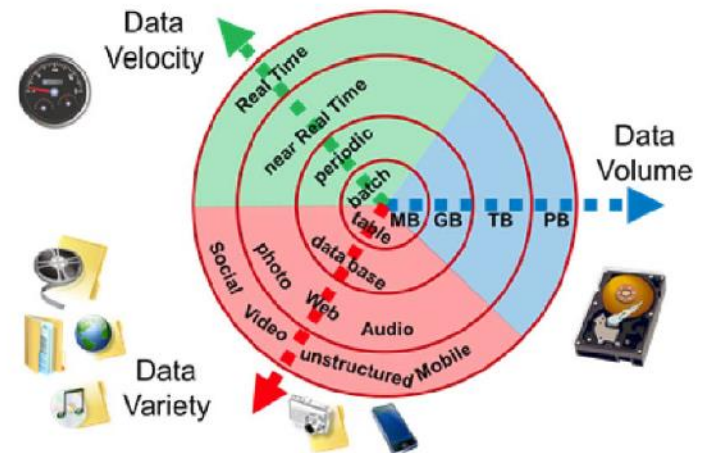
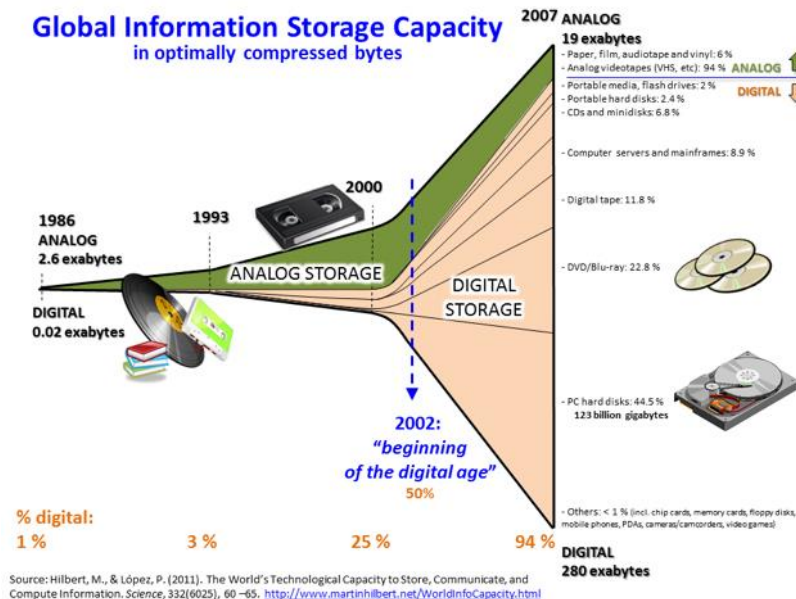
- **Analizom finansijskih podataka** koji se danas generišu velikom brzinom može se doći do **novih znanja** na osnovu kojih se mogu donositi **bolje odluke**, želja je imati **pravu informaciju u realnom vremenu**
- **Paralelizacija** donosi unapređenja kod algoritama u oblastima kao što su **određivanje cene finansijskih instrumenata, upravljanje rizicima i simulacija tržišta**
- Za rešavanje problema u oblasti ekonomije i finansija, danas se koristi **kombinacija heterogenog paralelnog i distribuiranog računarstva**



Izvor: <https://www.fool.com.au/2020/02/03/oil-search-share-price-drops-7-on-png-decision/>

Obrada velikih skupova podataka

- Količina podataka koja se čuva u svetu duplira se otprilike na svake dve godine, veličina „digitalnog univerzuma“ u 2020. je oko 44 ZB (10^{21} bajta), ako nisu predmet analize, skoro svi ti podaci su beskorisni
- Oblasti nauke o podacima i sistema za rad sa velikim količinama podataka se u velikoj meri oslanjaju na paralelnu i distribuiranu obradu



Izvor: https://en.wikipedia.org/wiki/Big_data

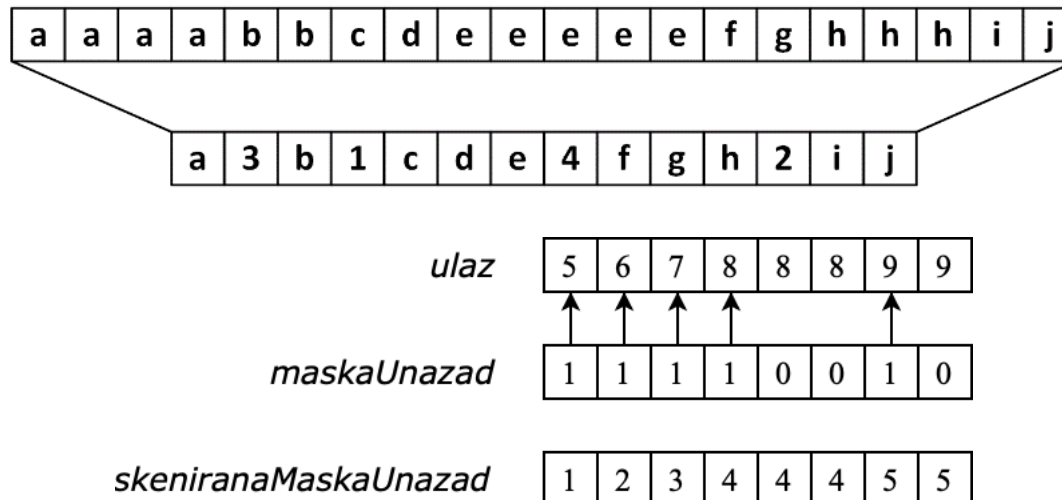
Obrada velikih skupova podataka

- Često korišćena strategija za obradu velikih skupova podataka podelom na delove koji se mogu paralelno i distribuirano obrađivati jeste obrazac **mapiranje/redukcija**
- Prvo se vrši **mapiranje** (filtriranje, sortiranje), potom ide **redukcija** (sumiranje), **stablo redukcije** (engl. *reduction tree*) se koristi kako bi se sumirali rezultati za svaki od delova u konačni odgovor
- **MapReduce** je programski model koji predstavlja primer implementacije ovog obrasca – open source projekat Apache Hadoop
- Na GPU, svaka nit može da obradi deo podataka i pritom koristi deljenu memoriju za čuvanje međurezultata, brzina rada **atomičnih operacija**, koje su ključne za **efikasnu redukciju**, je značajno unapređena u novijim GPU arhitekturama



Primer – Kompresija podataka

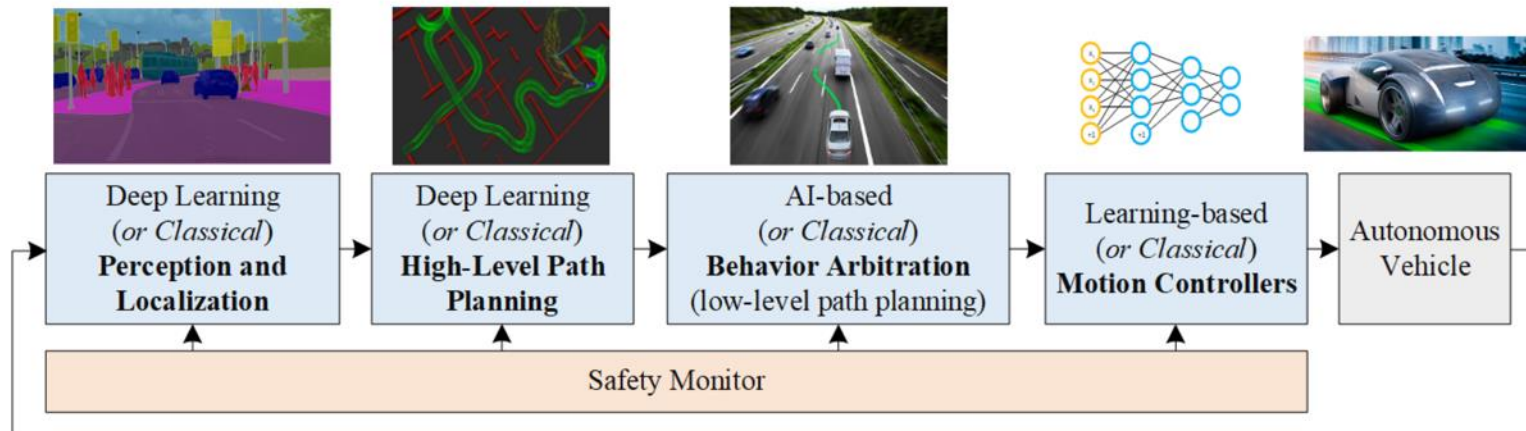
- Kompresiona tehnika **ran-lengt** (engl. *run-length encoding* – *RLE*) može se efikasno paralelizovati primenom obrasca **skeniranja** (engl. *scan*) – PARLE algoritam
- Skeniranjem se dobijaju izlazne pozicije svih znakova i time omogućava paralelizacija do nivoa da svaki od elemenata računa zasebna nit, veoma pogodna organizacija izračunavanja za visoko-paralelnu GPU arhitekturu



Izvor: Dragan, Gajić, “Uvod u kompresiju podataka – osnovni pojmovi i entropijske kompresione tehnike”

Mašinsko učenje

- **Duboko učenje** (engl. *deep learning*) zasniva se na visoko-paralelnoj obradi na grafičkim procesorima
- **Duboke neuronske mreže** imaju veći broj slojeva između ulaznog i izlaznog (rekurentne, konvolucione, transformeri...)
- **Primene:** prepoznavanje slike i govora, bioinformatika, obrada jezika, sistemi za preporučivanje, autonomna vožnja (Nvidia DRIVE OS, DriveWorks SDK)



Izvor: <https://arxiv.org/pdf/1910.07738.pdf>

Primer – Duboko učenje na Nvidia DGX



POWERED BY 4 NVIDIA TESLA V100 GPUs
BUILT ON THE LATEST NVIDIA VOLTA™ GPU ARCHITECTURE



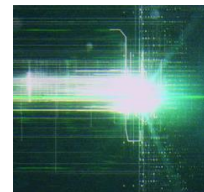
The power of
500 TFLOPS of AI Power



Get started
within one hour



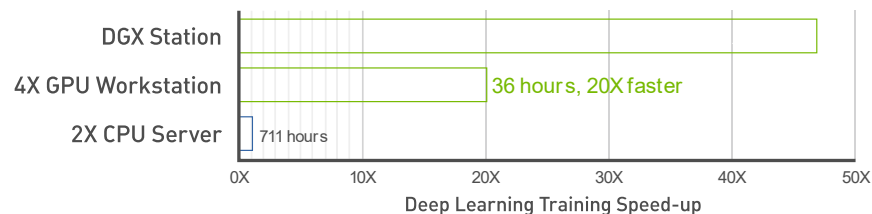
No data center?
No problem.



Access to
AI expertise

ITERATE AND INNOVATE FASTER
UNPARALLELED DEEP LEARNING TRAINING PERFORMANCE

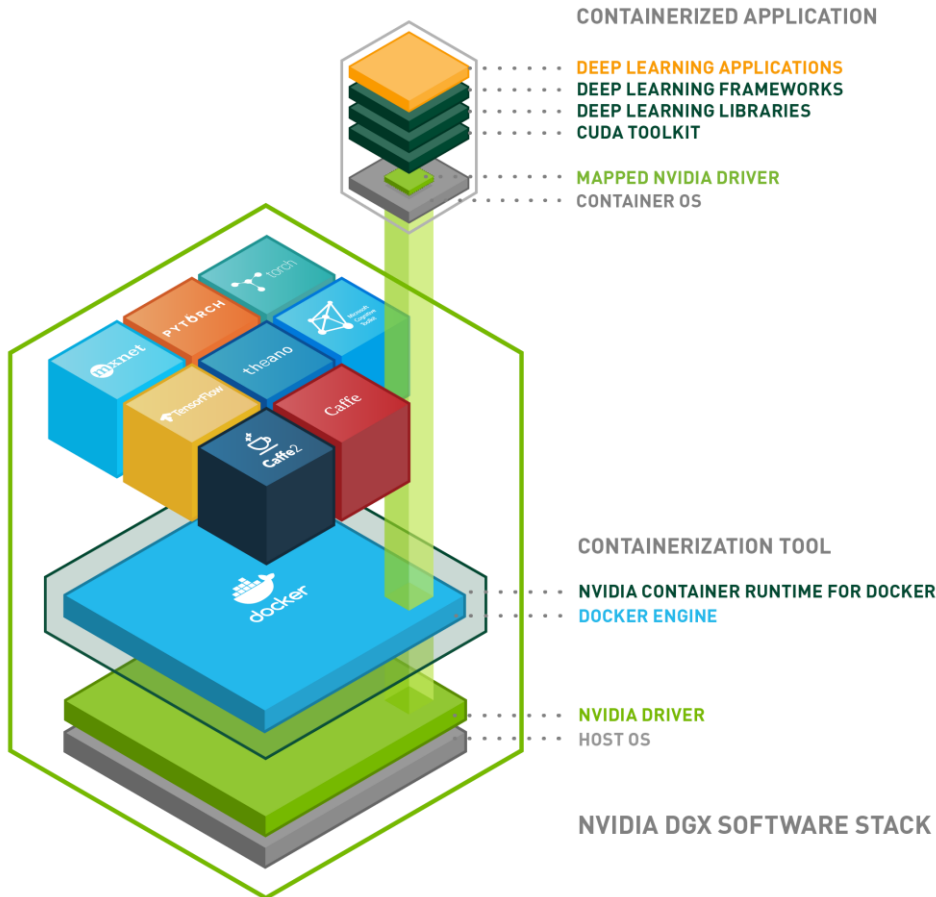
NVIDIA DGX Station Delivers 47X Faster Training



DGX Station performance projected based on DGX (with Tesla V100) Workload: ResNet50, 90 epochs to solution | CPU Server: Dual Xeon E5-2699 v4, 2.6 GHz. Projections subject to change.

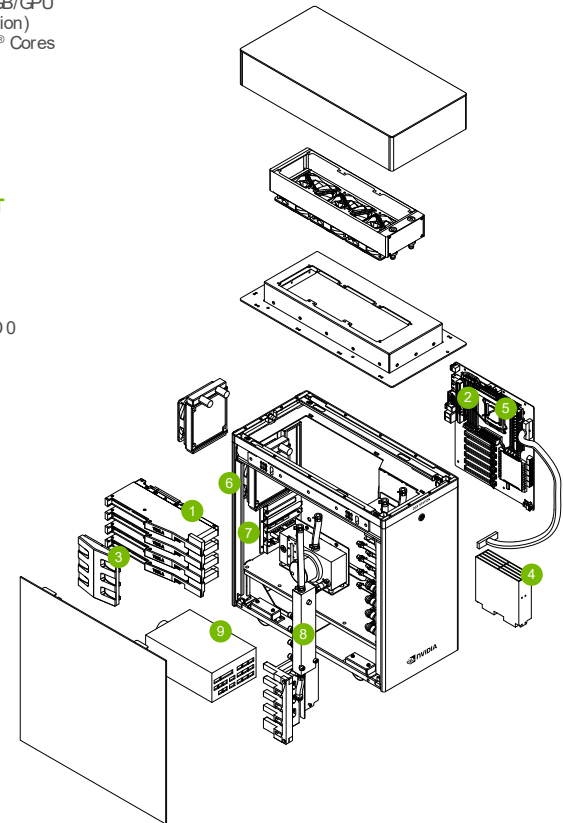
Primer – Duboko učenje na Nvidia DGX

Softver



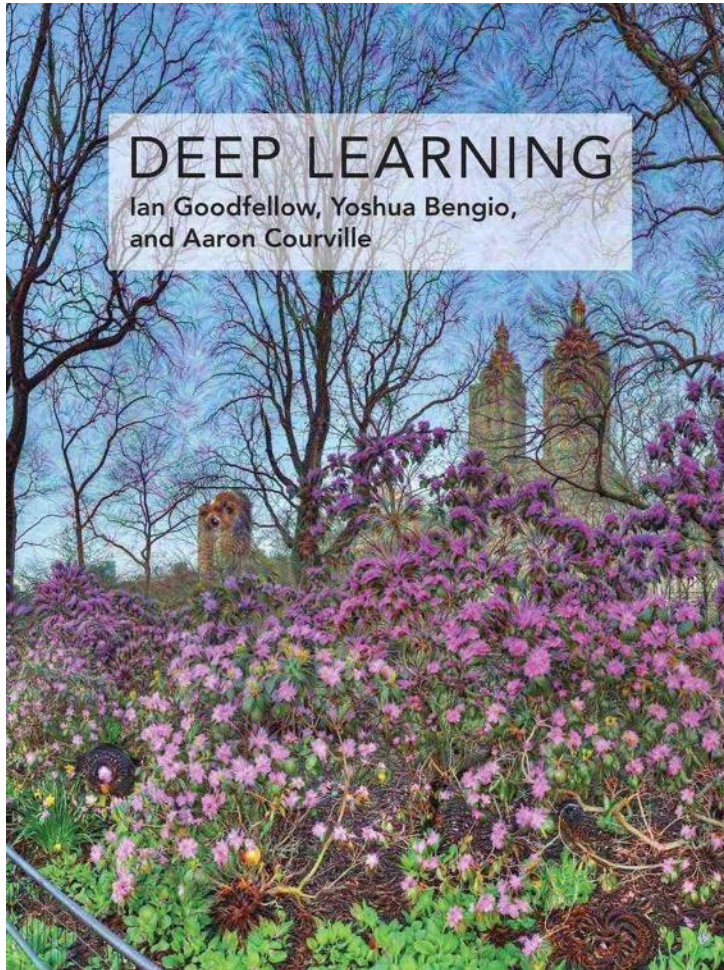
Hardver

- 1. GPUS**
4X NVIDIA Tesla® V100 32 GB/GPU
500 TFLOPS (Mixed Precision)
20,480 Total NVIDIA CUDA® Cores
2,560 Tensor Cores
- 2. SYSTEM MEMORY**
256 GB RDIMM DDR4
- 3. GPU INTERCONNECT**
NVIDIA NVLink™,
Fully Connected 4-Way
- 4. STORAGE**
Data: 3 x 1.92 TB SSD RAID 0
OS: 1 x 1.92 TB SSD
- 5. CPU**
Intel Xeon E5-2698 v4
2.2 GHz 20-Core
- 6. NETWORKING**
2X 10 GbE
- 7. DISPLAYS**
3X DisplayPort,
4K Resolution
- 8. COOLING**
Water-Cooled
- 9. POWER**
1500 W

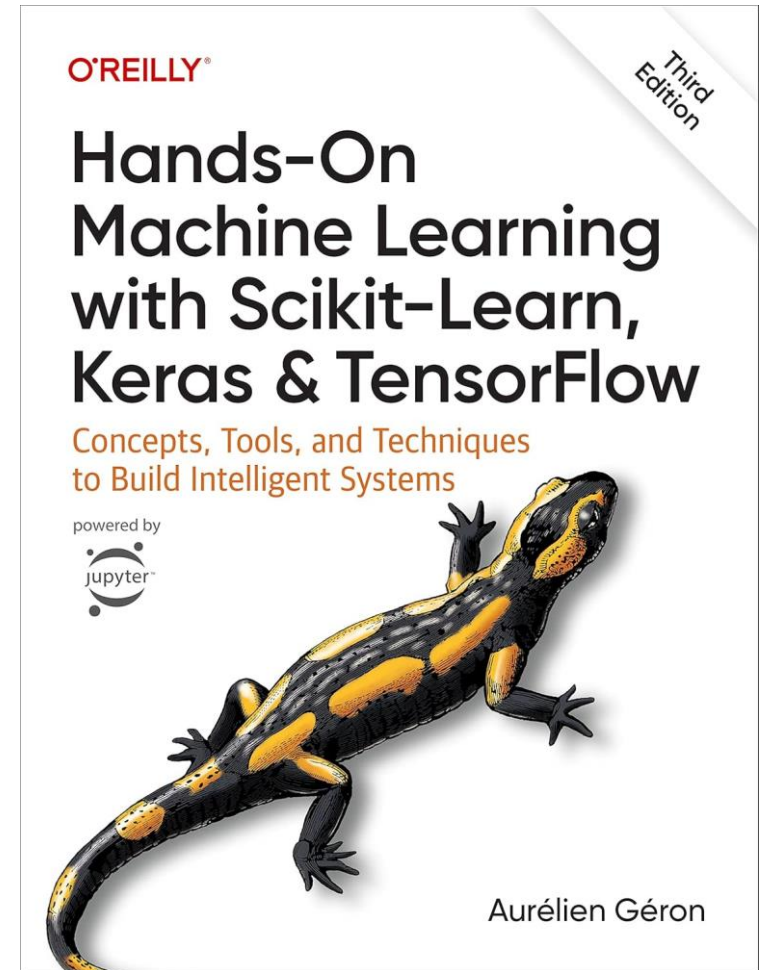


Izvor: <https://www.nvidia.com/content/dam/en-zz/Solutions/Data-Center/dgx-station/dgx-station-print-Infographic-738375-nvidia-web.pdf>

Literatura – Duboko učenje



<https://www.deeplearningbook.org>



<https://www.oreilly.com/library/view/hands-on-machine-learning/9781098125967/>